NORTHWESTERN UNIVERSITY | SCHOOL OF PROFESSIONAL STUDIES

PREDICT 402: Introduction to Predictive Analytics and Data Collection Winter 2016

Gordon Swartz, Instructor (DBA, Harvard Business School)

gordon.swartz@northwestern.edu

Office hours are on Tuesdays, with appointments between 7:00PM – 9:30PM CST, during weeks when we do not have sync sessions. As necessary, I will also accommodate phone / web-meeting appointments at other times, provided that we agree a mutually workable meeting time. Please ensure that all appointments are confirmed.

Every one of you is encouraged to reach out to me, if you are lagging in covering the course topics or if you anticipate a change in your circumstances that will affect your completion of the course.

Course Description

This course introduces the field of predictive analytics, which combines business strategy, information technology, and modeling methods. The course reviews the benefits and opportunities of data science, organizational and implementation issues, ethical, regulatory, and compliance issues. It discusses business problems and solutions regarding traditional and contemporary data management systems and the selection of appropriate tools for data collection and analysis. It reviews approaches to business research, sampling, and survey design.

Time is spent interpreting performance-based organizational issues while concurrently identifying solutions and identifying best practices to plan for engaging, implementing, and sustaining organizational change.

Texts – Required

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Boston, MA: Harvard Business School. [ISBN-13: 978-1422103326]

Few, Stephen (2013). *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring*, Oakland, Calif.: Analytics Press. [ISBN-13: 978-1938377006]

Laursen, G. & Thorlund, J. (2010), *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*, Wiley. [ISBN: 978-0-470-89061-5]

Groves, R.M., Fowler, F. Jr., Couper, M.P., Lepkowski, J.M., Singer E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, NJ: Wiley. [ISBN-13: 978-0470465462]

Texts – Recommended (Books you may use in other courses)

Thomas Miller (2015), *Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science*, Pearson Publications, ISBN-13: 978-0-13-389211-6

Texts – Recommended

Franks, B. (2012). *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*. Hoboken, NJ: Wiley. [ISBN-13: 978-1-118-20878-6]

Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.* Hoboken, NJ: Wiley. [ISBN-13: 978-1118356852]

Video

Moneyball, the 2011 movie, starring Brad Pitt. This movie should be available in the library resources.

Software

None required. The SAS product **JMP Pro** is available for downloading freely as part of the course offering. JMP Pro can be used for graphical visualizations of data and may be a useful tool for the dashboard assignment. (These assignments may also be completed with your choice of graphics / presentation tools, e.g., Microsoft Powerpoint or Apple's Macintosh Keynote software.)

Prerequisites

None.

Learning Goals

The goals of this course are to:

- Articulate the value of predictive analytics as a core business strategy and ways to implement analytics capabilities in an organization
- Know how to judge the analytical maturity level of an organization
- Transform business plans into analytics supported plans
- Analyze and evaluate traditional data management, big data management and analytics infrastructures
- Map internal and external business opportunities and various analytical methods associated with them
- Select the appropriate analytic approach given a business situation
- Understand the challenges of building analytics teams, managing analytics teams and their performance
- Articulate various measures of performance and design powerful dashboards
- Identify various third party data sources, constraints in using data such as data quality issues, and missing value problems for analytical applications
- Evaluate sample survey methods and create sample survey proposals for collecting data
- Formulate solutions for resolving data quality issues

Evaluation

The student's final grade will be determined as follows:

- Session 3: "Moneyball" Analysis and Identifying Big Data and Analytics Applications (***100 points***)
- Session 4: Case Study Collection Theme (1-2 paragraphs, ***feedback only***)
- Session 5: Dashboard and Executive Summary (***100 points***)
- Session 7: Case Study Collection (***150 points***)
- Session 8: Survey Design and Implementation Topic (1-2 paragraphs, ***feedback only***)
- Session 10: Survey Design and Implementation (***150 points***)
- Sessions 1–10: Discussion Board Participation (***100 points***, 10 points per session)

Total Points: 600 pts.

Each student may earn "bonus points" from one additional assignment, as follows:

- Communicating analytical insights: analysis and critique of: a) slide (***10 points***) or b) dashboard (***20 points***)

Grading Scale

A = 93%–100% (558–600 points)
A- = 90%–92% (540–557 points)
B+ = 87%–89% (522–539 points)
B = 83%–86% (498–521 points)
B- = 80%–82% (480–497 points)
C+ = 77%–79% (462–479 points)
C = 73%–76% (438–461 points)

C- = 70%–72% (420–437 points)
F = 0%–69% (0–419 points)

Discussion Board Etiquette

The purpose of the discussion boards is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. Active and frequent participation is encouraged. It is important to keep pace with the discussion and to ensure that your comments are relevant to the current topics and enriching to the discussion.

Cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency is not unimportant, but content of the message is paramount. Please remember to cite all sources—when relevant—in order to avoid plagiarism.

The deadline for posting to each week's discussion forum is Sunday, 11:55 p.m. (Central Time). However, please note that the discussion value and discussion quality of posts that are made near or at the deadline is naturally limited.

Important guidelines:

Post your viewpoints first and then discuss others' viewpoints. Good rules of thumb to gain full credit of 10 points each week: 1) post at least one substantive comment each week well in advance of the deadline, 2) read and reply to others' comments with extensions and enhancements as well as thoughtful, respectful critiques. (The quality of your posting and how others value your postings is important.)

- Single statements, mostly implying "I agree" or "I do not agree" are not counted as substantive postings, but are useful in providing feedback to others.
- Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience.
- Continued low-quality, tweet-type postings will get negative points.

Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. To measure class participation (or attendance), your participation in threaded discussion boards is required, graded, and paramount to your success in this class. Please note that any scheduled synchronous or "live" meetings are considered supplemental and optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation.

Late Work

Unless otherwise noted, all work is due on the assigned day by 11:55 p.m. (Central Time). This includes all graded assignments and participation in the discussions. Late work is not accepted.

One more piece of advice—do not fall behind. We cover a lot of material in this course, and falling behind is the primary reason why folks fail. To that end, you have below the due dates for the entire course. It is much, much better to be ahead than behind.

Academic Integrity at Northwestern

Students are required to comply with University regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor before the assignment is due and/or examine the University Web site. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit <www.scs.northwestern.edu/student/issues/academic_integrity.cfm>.

Plagiarism is one form of academic dishonesty. Students can familiarize themselves with the definition and examples of plagiarism, by visiting <www.northwestern.edu/uacc/plagiar.html>. A myriad of other sources can be found online.

Some assignments in this course may be required to be submitted through Turnitin, a plagiarism detection and education tool. You can find an explanation of the tool on the Canvas learning platform. In brief, Turnitin compares the submitted assignment to millions of documents in large databases. It then generates a report showing the extent to which text within a paper is similar to pre-existing sources. The user can see how or whether the flagged text is appropriately cited. Turnitin also returns a percentage score, indicating the percentage of the submitted paper that is similar or identical to pre-existing sources. High scores are not necessarily bad, nor do they necessarily indicate plagiarism, since the score does not take into account how or whether material is cited. If a paper consisted of one long quote that was cited appropriately, it would score 100%. This would not be plagiarism, due to the appropriate citation. However, submitting one long quote would probably be a poor paper. Low scores are not necessarily good, nor do they necessarily indicate a lack of plagiarism. If a 50-page paper contained all original material, except for one short quote that was not cited, it might score around 1%. But, not citing a quotation is still plagiarism.

Turnitin includes an option in which the student can submit a paper and see the resultant report before submitting a final copy to the instructor. This ideally will help students better understand and avoid plagiarism.

Other Processes and Policies

Please refer to your SPS student handbook for information about program processes and policies: <http://www.scs.northwestern.edu/program-areas/graduate/student-handbook.php>

A quick, online reference to formatting citations in American Psychological Association (APA) style is: APA Citation Guide (7 September 2015), Referenced from http://www.bibme.org/citation-guide/apa/

Course Schedule

Important Note: Changes may occur to the syllabus at the instructor's discretion.
When changes are made, students will be notified via an announcement on the learning platform.

Session 1 – Success Stories of Analytics Competition

Learning Objectives

After this session, the student will be able to:

- Define analytics
- State the importance of analytics in business competition
- Identify ethical and legal issues of using analytics
- Assess common attributes of analytically competitive businesses
- Rank the stages of analytic competition
- Clarify importance of dashboards
- Identify common mistakes in dashboard design

Course Content

Textbook Reading

Competing on Analytics, Chapters 1–3

Information Dashboard Design, Chapters 1–3

Course Reserves, HBR or Internet sourced articles.

Davenport, Thomas H., (2013) Analytics 3.0: in the new era, big data will power consumer products and services (Spotlight on Making Your Company Data-Friendly). *Harvard Business Review*. Dec. 2013, Vol. 91(12), p. 64(8)

- Search in the web for definition of Web 3.0 and/or IT 3.0.
- How well do these definitions correspond to the same level of organizational maturity? (Why should we expect these definitions to correspond to the same level of maturity?)
- How well does the Analytics 3.0 definition apply to traditional data (as opposed to "big data")?

Barton, Dominic; Court, David., (2012) Making advanced analytics work for you: a practical guide to capitalizing on big data. (Spotlight on Big Data)(Cover story). *Harvard Business Review*. Oct. 2012, Vol. 90(10), p.78(6)

- How do organizations prepare for big data and advanced analytics? How does this help or hinder efforts to "compete on analytics"?

Optional reading: *Predictive analytics,* Preface, Introduction, Appendices A and B.

Movie Viewing

Watch the *Moneyball* movie, with critical observation of individual and organizational behaviors, team dynamics and change management.

Pitt, B. (Actor, Producer), & Miller, B. (Director), (2011), *Moneyball* [Motion Picture], United States: Columbia Pictures.

Lecture Videos

- What Is Analytics and Why Is It Important?
- Strategic Data Assets, Tools, and Big Data
- View the tutorials on "Basics of APA Style of Writing"
- Listen to the audio on six parts to a well structured write up

Discussion Board

Each session you are required to participate in the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topic(s), visit the discussion board in Canvas. The due date and time for posting to each week's discussion forum is Sunday at 11:55 p.m. (Central Time).

Assignments

None.

Sync Session

SYNC SESSION: Tuesday evening, January 5, 2016 from 7:00 to 9:30 p.m. (Central Time).

Session 2 – Industry Surveys, Big Data Cases, BI and Analytical "Maturity"

Learning Objectives

After this session, the student will be able to:

- Assess the analytic capabilities of an organization
- Distinguish the stages of organizational analytic competition
- Evaluate the way organizations navigate the stages of becoming an analytic competitor
- Compare the roles played by analytic executives, analytic professionals, and analytic amateurs
- Review the literature on big data

Course Content

Textbook Reading

Competing on Analytics, Chapters 4 – 6

Information Dashboard Design, Chapters 4 and 5

Optional reading: *Predictive analytics,* Student-selected chapters.

Course Reserves, HBR or Internet sourced articles.

Kiron, D., Prentice, P.K., Boucher-Ferguson, R. (May 12, 2014). The analytics mandate. *Global Executive Study and Research Project. MIT Sloan Management Review*
(Retrieved from http://sloanreview.mit.edu/projects/analytics-mandate/)

- Many companies that use analytics are successful. So, what factors impede organizations from becoming analytically competitive?

Marco, O, & Lakhani, K. (Nov 2014) Digital ubiquity: How connections, sensors, and data are revolutionizing business. *Harvard Business Review*, Vol. 92(11), pp. 90-99.

- What new business processor opportunities does "digital ubiquity" imply for your organization?

Lecture Video

- BI Part 1 - Analytics among High vs. Low Performing Cos. and Why Analytics is Unique
- BI Part 2 - Business Intelligence, Competition, and Maturity Levels
- Top Videos on Learning BIG Data Science (http://tiny.cc/LBDS)

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

None.

Sync Session

None.

Session 3 – Internal and External Processes and Analytical Methods

Learning Objectives

After this session, the student will be able to:

- Identify analytic techniques used to analyze internal business processes
- Select the appropriate analytic applications for a given internal business process
- Identify analytic techniques used to analyze external business processes
- Select the appropriate analytic applications for a given external business process
- Specific dashboard design practices to tap into visual perception
- Integrate the six elements of business intelligence architecture
- Specify the relationship between the six elements of business intelligence architecture

Course Content

Textbook Reading

Competing on Analytics, Chapters 7 – 9

Information Dashboard Design, Chapters 6 – 9

Business Analytics for Managers, Chapters 1 – 2

Optional reading: *Predictive analytics,* Student-selected chapters.

Optional reading / skim: *Taming the Big Data Tidal Wave,* Chapters 7–8

Course Reserves, HBR or Internet sourced articles.

Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (07 November 2011) Analytics: The widening divide. How companies are achieving competitive advantage through analytics. *MIT Sloan Management Review.* Retrieved from http://sloanreview.mit.edu/reports/analytics-advantage/ (http://tiny.cc/widening-divide)

- How do the three case studies differ in terms of their business models and customer behaviors? Which one most resembles your organization or your favorite industry?

Lecture Video

- Analytical Methods for Internal and External Processes

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

Moneyball Analysis and Identifying Big Data and Analytics Applications paper is **due Sunday, January 24, 2016 at 11:55 p.m.** (Central Time). For more information, refer to the "Assignments" document on Canvas.

Sync Session

None.

Session 4 – Different Analytics Levels and the Strategic / Operational Value of Dashboards

Learning Objectives

After this session, the student will be able to:

- Organize the components of the business analytics model
- Assess the role of data in the business analytics model
- Classify the different types of links between business analytics and strategy
- Recognize the types of analytic information available to inform the three disciplines outlined
- Identify design considerations for dashboards

Course Content

Textbook Reading

Business Analytics for Managers, Chapters 3 – 9

Information Dashboard Design, Chapters 10 – 11

Optional reading: *Taming the Big Data Tidal Wave,* Chapters 1–3

Course Reserves, HBR or Internet sourced articles.

Chaudhuri, S., Dayal, U., & Narasayya, V. (August 2011). An overview of business intelligence technology.*(Review Articles)(Report). Communications of the ACM*, Vol. 54, pp. 88-99. Referenced from: http://cacm.acm.org/magazines/2011/8/114953-an-overview-of-business-intelligence-technology/fulltext

- Explain Fig. 1 and Fig. 4 and the concepts of RDBMS and CEP. Why are web log data the common denominator in CEPs?
- Explain dimensional data and star schema in traditional data management.

Harris, J.G., & Mehrotra, V. (Fall 2014) Getting value from your data scientists. *(Intelligence / Data and Analytics). MIT Sloan Management Review*, Vol 56, Issue #1, pp 15-20. (http://tiny.cc/getting-value)

- Is getting value from data scientists the same as getting value from data? Are the differences between analysts and data scientists natural to the life cycle of their activities? Why might these types of distinctions need to be raised in managing analysts and data scientists?

Lecture Video

- Strategic Metrics, KPIs, KLIs, and best practices for Dashboards

Discussion Board

Each session you are required to participate in the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topic(s), visit the discussion board in Canvas. The due date and time for posting to the discussion for this session is Sunday at 11:55 p.m. (Central Time).

Assignments

Case Study Collection Theme draft – for feedback only – is **due Sunday, January 31, 2016 at 11:55 p.m**. (Central Time). For more information, refer to the "Assignments" document on Canvas.

Sync Session

None.

Session 5 – Information Strategy, Designing Dashboards and Privacy

Learning Objectives

After this session, the student will be able to:

- Compare and contrast lag and lead information
- Distinguish how lead versus lag information can be used in the development and management of a new business process
- Distinguish how lead versus lag information can be used to optimize existing processes
- Classify key performance indicators into their suggested business functions
- Apply a strategy mapping process to match analytic techniques to information requirements
- Explain the difference between data, information, and knowledge
- Evaluate the importance of each of the analyst competencies
- Evaluate the advantages and disadvantages of different types of analytic reports
- Formulate business examples of when the use of data-driven versus data mining versus explorative analytic methods would be appropriate
- Compose effective business requirement documents

Course Content

Textbook Reading

Information Dashboard Design, Chapters 12 – 14

Optional reading: *Taming the Big Data Tidal Wave,* Chapters 4–6

Course Reserves, HBR or Internet sourced articles.

"AAA Funds: The KnowledgeSCORE Lead Optimization Pilot" case study.

Berinato, Scott (2014). With big data comes big responsibility: an interview with MIT Media Lab's Alex "Sandy" Pentland. *Harvard Business Review*, Vol. 92(11), pp. 100 – 109.

- Who owns the data? Why do consumers part with their personal data? Will the value of data increase significantly if consumers are allowed to share in the revenue of selling such data? Should and could privacy be an absolute right?

Lecture Video

- Development and Deployment of an Information Strategy

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

AAA Funds: Executive Summary and Dashboard is **due Sunday, February 7, 2016 at 11:55 p.m.** (Central Time). For more information, refer to the "Assignments" document on Canvas.

Sync Session

None.

Session 6 – Survey Methodology – Target Populations, Sampling Frames and Sampling Errors

Learning Objectives

After this session, the student will be able to:

- Explain the importance of sampling in analytics
- Compare and contrast different sampling techniques

Course Content

Textbook Reading

Survey Methodology, Chapters 1–4

Optional reading: *Taming the Big Data Tidal Wave,* Chapter 7

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

None.

Sync Session

SYNC SESSION: Wednesday evening, February 10, 2016 from 7:00 to 9:30 p.m. (Central Time).

Session 7 – Collecting Survey Data and How to Avoid Errors and Biases

Learning Objectives

After this session, the student will be able to:

- Evaluate the benefits and limitations of data collection methodologies
- Evaluate the benefits and limitations of data collection modalities
- Apply the fundamentals of survey design to develop an effective survey
- Explain the relationship between components in a data warehouse
- Identify business systems that may generate data
- Organize the steps in the extraction, transformation, loading (ETL) process
- Identify and anticipate potential sources of poor quality data
- Evaluate the effects of poor quality data
- Identify potential sources of data in an organization
- Assess the relationship between the usability and the availability of data

Course Content

Textbook Reading

Survey Methodology, Chapters 5 – 7

Optional reading: *Taming the Big Data Tidal Wave,* Chapters 10–11

Course Reserves, HBR or Internet sourced articles.

Ross, B. (2009). Ten tips for winning at consumer centricity - for retailers and manufacturers. *Journal of Consumer Marketing.* Vol. 26, No. 6, pp. 450 – 454.

- Which of these tips are usable or adaptable to financial services, health care or another industry that interests you? Why / why not?

Lecture Video

- Survey Methodology – Part 1 and Part II

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

Case Study Collection Final Assignment is **due Sunday, February 21, 2016 at 11:55 p.m.** (Central Time). For more information, refer to the "Assignments" document on Canvas.

Sync Session

None.

Session 8 – Handling Missing Values, Imputation (It really works!)

Learning Objectives

After this session, the student will be able to:

- Assess the impact of missing data on the analytic process
- Appraise the benefits and limitations of different data imputation techniques

Course Content

Textbook Reading

Survey Methodology, Chapters 8 – 10

Lecture Video

- Sample size, Neyman Allocation, and Minimum Allocation

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

Survey Design and Implementation Topic draft – for feedback only – is **due Sunday, February 28, 2016 at 11:55 p.m.** (Central Time). For more information, refer to the "Assignments" document on Canvas.

Sync Session

SYNC SESSION: Tuesday evening, February 23, 2016, from 7 to 9 p.m. (Central Time).

Session 9 – Data Quality, Data Integration, Predictive Modeling and Scoring

Learning Objectives

After this session, the student will be able to:

- Explain the importance of sampling in big data environments
- Detail effective methodologies for gathering data, ensuring data quality, specifying relevant metrics and models and reporting impactful, credible insights

Course Content

Textbook Reading

Survey Methodology, Chapters 11 – 12

Internet Sourced Materials

Sambamoorthi, N. (2014). Systematic approach to data requirements and data collection. (http://tiny.cc/analysis-business)

Lecture Video

- Relationship among Issue Tree Method, Data Collection, and Analytics

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

None.

Sync Session

None.

Session 10 – Course Wrap-Up

Learning Objectives

- Earlier course topics are integrated and applied.

Course Content

No new required readings.

Discussion Board

As noted above, every student is expected to participate in the discussion board every week. The final due date for posting to the discussion for each weekly session is Sunday at 11:55 p.m. (Central Time).

Assignments

Survey Design and Implementation assignment is **due Sunday, March 13, 2016 at 11:55 p.m.** (Central Time). For more information, click Assignments on the left navigation panel in Canvas, and scroll to this assignment's item.

Sync Session

None.

Calendar: Predict 402, Winter Term 2016

Predict 402	**Winter 2016**
Start of Term **- Sync Session #1, Tuesday evening, Jan.5**	1/4/2016
End of Session 1	1/10/2016
End of Session 2	1/17/2016
End of Session 3 - Monday, Jan. 18th is Martin Luther King Holiday **- Moneyball assignment due**	1/24/2016
End of Session 4 - **Draft** – 1-2 paragraphs on case theme due	1/31/2016
End of Session 5 **- AAAF exec sum / dashboard assignment due**	2/7/2016
End of Session 6 **- Sync Session #2, Wednesday evening, Feb. 10**	2/14/2016
End of Session 7 **- Case Study Theme assignment due**	2/21/2016
End of Session 8 **- Sync Session #3, Tuesday evening, Feb. 23** - **Draft** – 1-2 paragraphs on survey assignment due	2/28/2016
End of Session 9	3/6/2016
End of Session 10 **- Survey design and implementation assignment due**	3/13/2016
Grades Due in CAESAR	**3/21/2016 at 3pm**